Training and timing local scalar enrichments under global pragmatic pressures

Citation for published version:
Chemla, E, Cummins, C & Singh, R 2016, 'Training and timing local scalar enrichments under global pragmatic pressures', *Journal of Semantics*, vol. 34, no. 1, pp. 107-126. https://doi.org/10.1093/jos/ffw006

Digital Object Identifier (DOI):
10.1093/jos/ffw006

Training and timing local scalar enrichments under global pragmatic pressures*

Emmanuel Chemla[1], Chris Cummins[2], and Raj Singh[3]

[1] Laboratoire de Sciences Cognitives et Psycholinguistique (ENS, EHESS, CNRS), Département d'Etudes Cognitives, Ecole Normale Supérieure, PSL Research University
[2]University of Edinburgh, Department of Linguistics and English Language
[3]Carleton University, Institute of Cognitive Science

October 29, 2015

Abstract

Elementary sentences containing the quantificational determiner *some* seem to be ambiguous between a 'weak' existential meaning $\exists$ and a 'strengthened' *some but not all* meaning $\exists^+$. The strengthened meaning is commonly assumed to be the output of a general enrichment mechanism, call it $\mathcal{G}$ (for 'global'), that applies to the weak meaning of the sentence: $\mathcal{G}(\exists) = \exists^+$. The application of $\mathcal{G}$ has been shown to come with a processing cost (e.g., Bott and Noveck 2004). We used a self-paced reading task together with offline comprehension questions to investigate the interpretation of sentences containing *some* when embedded inside a disjunction, a position that $\mathcal{G}$ cannot access. Our findings suggest (i) that the strengthened meaning $\exists^+$ is available in embedded positions, suggesting that a mechanism of local strengthening $\mathcal{L}$ must be available: $\mathcal{L}(\exists) = \exists^+$, (ii) that local enrichment can be facilitated by global pragmatic pressures (Chierchia et al. 2008, Mayr and Romoli 2014), (iii) that subjects can be quickly trained to systematically prefer one of $\mathcal{G}$ or $\mathcal{L}$ to the other, (iv) that application of $\mathcal{L}$, like the application of $\mathcal{G}$, comes with a processing cost. We highlight consequences of our findings for debates about the characterization of enrichment mechanisms, focussing on the relation between $\mathcal{G}$ and $\mathcal{L}$.

*We thank Amir Anvari, Danny Fox, Lyn Frazier, Ted Gibson, Roni Katzir, Jacopo Romoli, Benjamin Spector, Ida Toivonen, Shravan Vasishth, Ken Wexler, and audiences at MIT, *XPrag 2013*, *AMLaP 2014*, and *TOM 6* at McGill University. The paper has improved thanks to helpful comments from three anonymous reviewers for *Journal of Semantics*, and the editor Bernhard Schwarz. The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013) / ERC Grant Agreement n.313610 and was supported by ANR-10-IDEX-0001-02 PSL* and ANR-10-LABX-0087 IEC, as well as by the Social Sciences and Humanities Research Council of Canada, grant number 435-2012-1573.

1 Introduction: Strengthening and processing

1.1 Global strengthening

Elementary sentences containing logical operators like *some* and *or* systematically generate two readings, a 'weak' and a 'strengthened' reading:

(1) The letter is connected to some of its circles.

a. Weak meaning: false if the letter is connected to none of its circles; otherwise true (= $\exists$)

b. Strengthened meaning: false if the letter is connected to none or all of its circles; otherwise true (= $\exists^+$)

(2) John ate pizza or apples.

a. Weak meaning: false if John ate neither pizza nor apples; true otherwise

b. Strengthened meaning: false if John ate neither pizza nor apples or if he ate both; true otherwise

The ambiguities in (1) and (2) appear cross-linguistically, and are never lexicalized (Horn 1972). For these and other reasons (e.g., McCawley 1981, Simons 2000, Sauerland 2012) the ambiguities are commonly explained by assuming that the lexical entries for *or* and *some* encode weak meanings which get strengthened by a general mechanism (Grice 1967). To a first approximation this mechanism can be thought of as a function that takes the utterance S and an alternative sentence S' and returns $\neg S'$. This negated alternative is the 'scalar implicature' of S, and the conjunction $S \wedge \neg S'$ is the 'strengthened meaning' of S, S^+.

(3) The letter is connected to some of its circles.

a. Weak meaning: $\exists$

b. Alternative: The letter is connected to all of its circles (= $\forall$)

c. Scalar implicature: $\neg\forall$

d. Strengthened meaning: $\exists \wedge \neg\forall$ (= $\exists^+$)

There are debates about the characterization of alternatives and about the strengthening mechanism itself. Most relevant to our discussion is the question whether the strengthening function is a domain-general reasoning mechanism or a domain-specific operator realized in the grammar.

On the domain-general view, $\exists^+$ is what a rational listener would conclude on the assumption that the speaker who uttered $\exists$ was obeying principles of rational cooperative social interaction that we assume are familiar (e.g., Grice 1967, Horn 1972, Gamut 1991, Spector 2005, 2006, Schulz and van Rooij 2006, van Rooij and Schulz 2004, Sauerland 2004, Russell 2006, Franke 2011). However, it has long been noted that domain-general principles must be supplemented with restrictions on scalar alternatives if the empirically attested implicatures are to be derived, and the restrictions do not themselves follow from domain-general considerations (e.g., Kroch 1972, Gazdar 1979, von Fintel and Heim 1999, Fox 2007a, Katzir 2007). Without such restrictions, Gricean reasoning does not not yield $\exists^+$ but rather ignorance inferences entailing that the speaker is ignorant about the stronger alternative $\forall$.[1] To overcome this limitation to ignorance inferences, all systems need to set grammatical restrictions on alternatives ('Horn scales').

Fox (2007a, 2013) suggests an alternative perspective that would allow domain-general reasoning to remain 'pure' (free of grammatical restrictions) by re-assigning grammatical stipulations to the grammar itself. Specifically, he suggests that a covert operator exh with a meaning like *only* is available in the grammar and that, when appended to $\exists$, the new sentence $exh(\exists)$ would have the same meaning as (4); without exh, pragmatic reasoning delivers ignorance inferences about the stronger alternative $\forall$ (cf. footnote 1):[2]

(4) The letter is connected to only some of its circles.
Literal Meaning: the letter is connected to some but not all of its circles (= $\exists^+$)

For the moment we need not take sides on the debate. To proceed we need to make the following assumptions, which so far as we can tell are common to both sides:

(5) Strengthening assumptions

a. Scalar items have a weak meaning (e.g., $[\![$*some*$]\!] = \exists$).
b. There is a strengthening function, $\mathcal{G}$, which strengthens the weak meaning of the asserted sentence (e.g., $\mathcal{G}(\exists) = \exists^+$). Crucially, we assume at this point that $\mathcal{G}$ only has access to the meaning of the entire sentence (it is 'global'); information about the meanings of subconstituents is lost at the root.[3]
c. Strengthening is an alternative-sensitive computation: $\mathcal{G}$ takes as input a sentence S, and a set of alternative sentences/propositions $ALT(S)$, and conjoins with $[[S]]$ the negation of some propositions in $ALT(S)$. (We provide a more specific characterization in later parts of the text.)

[1] We assume that the reasons for this are familiar. To briefly remind the reader, the main problem is that if alternatives are determined by relevance (answers to a question-under-discussion, e.g., Groenendijk and Stokhof 1984, Lewis 1988, Roberts 1996), then when $\forall$ is relevant, $\neg\forall$ will also be relevant because relevance in the intended sense is closed under negation – relevance is about *whether* a proposition is true. The alternatives $\forall$ and $\neg\forall$ are 'symmetric': each of them can be negated while maintaining consistency with $\exists$, but *not at the same time*. At best, one may negate that the speaker believes that any of them is true, which would yield that the speaker is ignorant about *whether* $\forall$ is true, but not that she holds the belief that it is false.

One may argue that this is not much of a problem if we take into account that $\neg\forall$ is not stronger than $\exists$, and hence might not play a role at all. But note that (a) non-stronger alternatives cannot easily be taken out of the picture and (b) the reasoning also applies if we replace $\neg\forall$ with $\exists \wedge \neg\forall$, which is plainly stronger than the initial sentence and ought to be an alternative too if relevance is closed under conjunction.

Horn-scales (Horn 1972) are used to break the symmetry by excluding $\neg\forall$ as an alternative. See especially the discussion of the 'symmetry problem' in Fox (2007a), Katzir (2007), Fox (2013), Fox and Katzir (2011).

[2] See also Chierchia (2004, 2006), Fox (2007a), Chierchia et al. (2008), Magri (2009b), Gajewski and Sharvit (2012), Fox (2013). Gazdar (1979) and Chemla (2009) also develop domain-specific systems, but these systems might be thought of as modules dedicated to conversational reasoning belonging to neither grammar nor central systems.

[3] Neo-Gricean proposals would identify $\mathcal{G}$ with the grammatically-restricted pragmatic reasoning laid out earlier, which naturally applies to whole sentences since it is a general purpose reasoning system that is supposed to process whole messages that may be received. The grammatical theory would identify $\mathcal{G}$ with matrix application of exh, which may in other occasions apply to smaller pieces of linguistic material, as we will shortly discuss. What is important is that under either approach $\mathcal{G}$ as we define it here does not have access to information about the meanings of sub-constituents.

d. Strengthening is optional, which we take to mean that application of $\mathcal{G}$ is optional (though see Magri 2009b, 2011).

1.2 Processing complexity

A persistent finding is that interpretation of a sentence containing *some* (henceforth simply *some*) with its strengthened meaning $\exists^+$ takes longer than when it is interpreted with its weak meaning $\exists$ (e.g., Noveck and Posada 2003, Bott and Noveck 2004, Breheny et al. 2006, Huang and Snedeker 2009, Bott et al. 2012, Chemla and Bott 2014; see Noveck and Reboul 2008, Katsos and Cummins 2010, Chemla and Singh 2014a,b for reviews). We will assume, then, that application of $\mathcal{G}$ to assertions of $\exists$ comes with some cost.

Our goal in this paper is to examine the processing of *some* when it occurs embedded in non-asserted positions, such as when *some* is embedded under disjunction: *the letter is connected to some or none of its circles*. Specifically, we will examine whether enrichments of *some* can be detected in such positions and, if so, whether they are costly like their global counterparts. Given the limitation of $\mathcal{G}$ to full sentences, it cannot apply inside the disjunction (neither disjunct is asserted). Nevertheless, there is evidence from offline judgments that local enrichments are available, suggesting that a mechanism of local strengthening $\mathcal{L}$ must exist.

1.3 Local strengthening

Our discussion has so far been limited to sentences containing a single logical operator. Call such sentences 'elementary sentences.'[4] What happens when scalar items are embedded under additional operators? Consider (6), which we assume has the logical form 'every letter x, x is connected to some of its circles' (we ignore here the likely quantifier movement out of object position):

(6) Every letter is connected to some of its circles.

By extending the general procedure in (3) from elementary sentences to sentences of arbitrary complexity, the constituent 'x is connected to some of its circles' would be converted to 'x is connected to all of its circles', generating the alternative (that gets pronounced as) *every letter is connected to all of its circles*. $\mathcal{G}$ would negate this alternative, producing the strengthened meaning that every letter is connected to some of its circles and that not every letter is connected to all of its circles. This does not, however, yield another reading that (6) has been argued to have, namely, that each letter is connected to only some of its circles (e.g., Chierchia 2004). To derive this embedded enrichment one might be tempted to posit an additional strengthening mechanism, $\mathcal{L}$, which would apply to the embedded occurrence of *some*

[4] Note that elementary sentences might be syntactically complex, containing multiple sentential constituents. For example, a disjunction $p \vee q$ contains the two disjuncts p and q, and a sentence containing a single quantificational noun phrase like *The letter is connected to some circles* plausibly has the logical form 'some circles x, the letter is connected to x', and thus contains at least one sentential sub-constituent ('the letter is connected to x'). On the 'formulas' view of the structure of quantificational noun phrases (Heim 1997), the LF for this sentence would contain even further propositional constituents (though see Kennedy 2014). We hope we can abstract away from these syntactic issues and focus solely on operator complexity in what follows. We thank a reviewer for helpful comments.

to produce the desired embedded $\exists^+$ meaning: *every letter* $x, \mathcal{L}($*x is connected to some of its circles*$)$.

It has been argued that the apparent embedded strengthening might actually be derivable through purely global reasoning. Specifically, it has been argued that the desired reading can be derived with $\mathcal{G}$ alone by (i) expanding the set of alternatives to also include *some letter is connected to all of its circles*, and (ii) allowing $\mathcal{G}$ to negate not only stronger alternatives but also those that are merely non-weaker than the assertion (Chemla 2009, Chemla and Spector 2011, Chemla and Singh 2014a,b). Thus, sentences like (6) do not provide evidence for $\mathcal{L}$; as noted earlier, the theory of strengthening needs to resolve many fine-grained choice-points about alternatives and the nature of strengthening, and (i) and (ii) are consistent with current knowledge (though see Fox 2007a, Note 35 for concerns with (i); we return to this in section 3.2).

A more direct motivation for $\mathcal{L}$ comes from the behaviour of *some* under disjunction. Consider the innocuous-looking sentence in (7):

(7) The letter is connected to some or all of its circles.

Without application of $\mathcal{L}$ at the first disjunct the sentence would actually violate 'Hurford's Constraint' (Hurford 1974), which bans disjunctions when one of the disjuncts entails the other (Fox 2006, Spector 2006, Fox 2007b, Chierchia et al. 2008; see also Gazdar 1979 and Simons 2000, and see Chemla 2009, Singh 2012, Meyer 2013, Katzir and Singh 2014 for attempts to derive the constraint from more general principles):[5]

(8) a. #John is an American or a New Yorker.
b. #John ate an apple or a fruit.

There are two things to note. First, the disjunctions in (8) are extremely odd. Second, (7) has no hint of oddness. If Hurford's Constraint is right, then (7) should pattern like (8), which it clearly does not.

Fox (2006), Spector (2006) and Chierchia et al. (2008) use this contrast to argue that a mechanism of local strengthening must be available at the first disjunct of (7). Without $\mathcal{L}$ the contrast between (7) and (8) remains unexplained, but the availability of $\mathcal{L}$ would allow (7) to be rescued by breaking the entailment between the disjuncts: 'the letter is connected to only some or all of its circles' would schematically be equivalent to '$\exists^+ \vee \forall$', and there is no entailment relation between $\exists^+$ and $\forall$. One may argue that $\mathcal{L}$ is allowed only in such special cases, when it comes to the rescue to satisfy an otherwise deviant sentence or it gives rise to special intonational markedness (e.g., Geurts 2009). However, as far we know there is no discussion of how this contrast could be explained using only global reasoning mechanisms.

[5]Chemla (2009) and Singh (2012) propose to derive the constraint from a contradiction between the inference that the speaker believes the assertion, $B(p \vee q)$, and the implicature that the speaker does not believe either disjunct (recall that for a Hurford disjunction $p \vee q$ is equivalent to one of its disjuncts). This does not extend to embeddings of Hurford Disjunctions: *John isn't a New Yorker or an American* is odd, but because $\neg(p \vee q)$ entails $\neg p \wedge \neg q$, there is no misleading ignorance implicature that can be generated. An alternative view relies on redundancy: a sentence is odd if one of its constituents could be deleted with no loss of information (Katzir and Singh 2014, Meyer 2013). This statement captures the oddness of matrix and embedded Hurford disjunctions, but it needs to be modified for reasons we don't discuss here. For our purposes, the important observation is that there is a contrast between (7) and (8) which can be described with Hurford's Constraint whatever the explanation behind the constraint.

Let us therefore assume for the moment that a local strengthening mechanism $\mathcal{L}$ exists. We want to find out whether embedded enrichments – applications of $\mathcal{L}$ – come with processing cost, like applications of $\mathcal{G}$ do (cf. section 1.2). In the classic Bott and Noveck (2004) paradigm, one compares the RTs of a sentence under its strengthened and unstrengthened readings, where the readings are determined by participants' truth-value judgments in contexts in which the different readings give different truth-values. Unfortunately, this method cannot be applied in sentences like (7) because the truth-conditions of the locally strengthened and unstrengthened sentences are the same: $\exists \vee \forall \iff \exists^{+} \vee \forall \iff \exists$. The argument for $\mathcal{L}$ in the case of (7) is based on patterns of felicity judgments, rather than truth-value judgments.

We would like to remain as methodologically conservative as possible, and thus would like to find cases of local strengthening that change global truth-conditions — in order to independently assess and manipulate parsing decisions. We would also like to preserve the key property of the Hurford paradigm, which *requires* that local strengthening takes place in order to satisfy a global pragmatic pressure (satisfaction of Hurford's Constraint) — this will allow us to derive fine-grained processing predictions at various points in time.[6] We saw that the purported embedded strengthening in (6) does yield a new reading, but as we noted earlier this datum might not provide evidence for $\mathcal{L}$. Furthermore, there are debates about how robust this reading is (e.g., Chierchia 2004, Sauerland 2004, Geurts and Pouscoulous 2009, Chemla and Spector 2011, Clifton and Dube 2010, Geurts and van Tiel 2013). Chierchia et al. (2008) produce sentences like (9) as examples of local strengthening that both obviate Hurford's Constraint and change global truth-conditions (note that, because it violates Hurford's Constraint, the weak meaning in (9-a) is actually unavailable in normal discourse):

(9) Of these ten problems, Jack solved the first and the second problem or he solved all of them.
   a. Weak meaning: Jack solved the first two problems, and possibly more (e.g., true if he solved only the first three problems)
   b. Locally strengthened meaning: Jack either solved only the first two problems, or he solved all of them (e.g., false if he solved only the first three problems)

In (7), strengthening of the first disjunct did not change global truth-conditions because, in a sense, the semantic effect of this strengthening is 'cancelled' by the second disjunct (the second disjunct reintroduces the proposition that is negated at the first disjunct). More precisely, the second disjunct *all* is the only alternative to *some*, and for all p and q the proposition $(p \wedge \neg q) \vee q$ is equivalent to $p \vee q$ which, in a Hurford configuration, is equivalent to p (when $q \subset p$).[7] Thus, there is no change in global truth-conditions. In (9), however, there are many more alternatives that can be negated that are intermediate in strength between the first and second disjunct, such as *Jack solved the first three problems*. This difference allows local strengthening to have global truth-conditional consequences: strengthening of the first disjunct negates these

[6] There is evidence that $\mathcal{L}$'s ability to rescue sentences from Hurford's Constraint is subject to incrementally evaluated constraints that block $\mathcal{L}$ from applying when the weaker disjunct is sentence-final. For example, in response to the question *who came to the party?*, A's answer is appropriate but B's is not: A: *(John or Mary) or both* vs B: # *Both John and Mary or (John or Mary)* (Singh 2008a,b, Fox and Spector 2008, 2015). This qualification is not relevant to our study, since the weaker disjunct is always initial.

[7] For ease of exposition, we will not carefully distinguish between sentences and propositions. However, when propositions are intended, $\wedge$ here should be understood as intersection, $\neg$ as set-complement, and $\vee$ as union.

alternatives, and the second disjunct cannot cancel these inferences. More abstractly, let p, q, r be propositions such that $p \subset r \subset q$. Then the proposition $(p \wedge \neg r \wedge \neg q) \vee q$ is not in general equivalent to $p \vee q$. In a Hurford configuration, the former is equivalent to $(p \wedge \neg r) \vee q$ while the latter is equivalent to p. Thus, local strengthening in such configurations – called 'distant entailing disjuncts' in Fox and Spector (2008) – can lead to global strengthening as well.

Example (9) thus satisfies our design considerations so far, but no study has investigated the processing complexity of enrichments in such configurations and, in fact, very little information exists about the processing profiles about the specific scalar enrichments involved in here to begin with. There is however evidence that different scalar items may pattern differently in terms of derivation rates (e.g., Reinhart 2006, Chemla 2013, van Tiel et al. in press), and therefore possibly on processing grounds, too (and one can find direct evidence of processing differences between *some* and numerals, e.g., in Huang and Snedeker 2009, Marty et al. 2013; see Chemla and Singh 2014b for a summary). We would thus like to stick with *some*, if possible, so that interpretation of any comparative findings might stand on firmer footing.

Our own experiment, discussed in detail in the next section, uses disjunctions like the following:

(10) The letter is connected to some or none of its circles.

On its literal meaning, this sentence provides no information: $\exists \vee \neg\exists$ is a tautology.[8] Pragmatic pressures to be informative in discourse (e.g., Grice 1967, Stalnaker 1978) might thus encourage subjects to apply $\mathcal{L}$ to the first disjunct, because such a reading would now convey the information that the letter is not connected to all of its circles: $\exists^+ \vee \neg\exists \iff \neg\forall$.

Disjunctions like (10) will be our critical items: they require local strengthening to satisfy a global pragmatic constraint, local strengthening produces a new meaning, and we use scalar items (*some*) whose processing profiles in matrix sentences are well-described. Our experiment, discussed in detail in the next section, first *trained* participants to apply one of $\mathcal{L}$ (the 'local' group) or $\mathcal{G}$ (the 'global' group) to items like (6),[9] and then presented them with critical items like (10). Assuming the training to be effective, the local group should tend to insert

[8] We considered the possibility that (10) could be a so-called *L-analytic* tautology in the sense of Gajewski (2004); that is, that it has the truth value 'true' independent of the non-logical lexical items involved. Concretely, all sentences of the form "The A is V to some or none of its B" are true, no matter what lexical items replace A, V and B. Such configurations are supposed to give rise to deviance judgments close to plain ungrammaticality judgments (e.g., von Fintel 1993, Gajewski 2004; see also Fox 2000, Fox and Hackl 2006, Abrusán 2014). However, as pointed by a reviewer, it is more plausible to assume that the form is "The A is V to some of its B_1 or none of its B_2" and even though the two Bs are realized in the same way at the surface, there is room to obtain a non-tautologous sentence with this structure, e.g., "The *letter* is *connected* to some of its *circles* or none of its *squares*." As a further argument against an L-analyticity approach to this sentence the reviewer rightly points out that plain L-analytic tautologies such as "Someone but John came to the party" are less acceptable than sentences of the form "The letter is connected to none, some but not all, or all of its circles". The difference in acceptability seems to be not merely a matter of degree, but also of kind; as noted earlier, L-analytic sentences are ungrammatical, while the tautologous sentence is perfectly grammatical but is an unhelpful speech act. Furthermore, there may be degrees of pragmatic acceptability; for example, Katzir and Singh (2014) note that Hurford sentences are less acceptable than tautologies. Note furthermore that the uninformative sentence 'none or some but not all or all' is similar in structure to our example of interest, but its tautologous aspect cannot be saved by potential local exhaustification.

[9] Danny Fox (p.c.), an anonymous reviewer, and Bernhard Schwarz point out that it is very important to consider whether our training and target sentences can receive their seemingly local readings through global strengthening (just like we showed that the $\neg\forall$ strengthened meaning of (6) can) and to carefully evaluate the consequences of such a

$\mathcal{L}$ at the first disjunct immediately. We expected this to have two detectable outcomes. First, there should be more $\neg\forall$ readings in the local group, because their initial analysis directly produces this reading whereas the global group generates it only after reanalysis. Second, both groups are expected to hit processing delays while reading the sentence, but the costs should be encountered at different regions: the local group should find difficulty *early* because they strengthen the first disjunct right away, and the global group should find difficulty *later* because they have to re-analyze at the second disjunct (to avoid tautology).

2 Experiment

2.1 Design and predictions

In our experiment, we examined the reading times for items such as (10). To avoid tautology, participants must interpret the 'some' of 'some or none' to mean 'some but not all'. That is, we expected these materials to elicit the application of $\mathcal{L}$. If $\mathcal{L}$ exists and the pragmatic pressure to be informative is real, participants should reject 'some or none' when 'all' is true; if not, they should accept 'some or none' when 'all' is true.

We also wished to investigate the time-course of this operation. $\mathcal{L}$ could operate either at the point at which 'some' is first encountered, or subsequently, as a consequence of reanalysis. We assume that $\mathcal{L}$, like $\mathcal{G}$, is optional in the positions in which it is licensed, and we investigated whether subjects could be trained to apply one of $\mathcal{L}$ or $\mathcal{G}$ in their initial analysis of the sentence.

To separate these possibilities, we proposed to compare the reading times for participants under two distinct training conditions intended to favor one or the other of these interpretation strategies. We biased one group of participants (the 'Local' group) to perform apparently local enrichments to embedded instances of 'some' under 'every' (e.g., *every letter is connected to some of its circles*) and another group (the 'Global' group) to perform only the global enrichments of these sentences. We then examined their self-paced reading times on disjunctive sentences like 'Letter D is connected to some or none of its circles.' The Global group should not perform immediate local enrichments. Consequently, they would need to reanalyse the sentences upon encountering 'or none' in order to avoid tautology. By contrast, the Local group may perform immediate local enrichments and thus may not need to go through a reanalysis. We therefore predicted that reading times for the Local group would be faster than for the Global group in the quantifier's spillover region 'of its circles.' We also expected to find that, if $\mathcal{L}$ incurs a cost, Local participants would be slower than Global participants in the quantifier region itself ('some or none').

2.2 Procedure

The experiment comprised 100 trials and was implemented using Ibex Farm.[10] In each trial, a sentence was presented one word at a time. Participants were instructed to press the space

possibility for the interpretation of our results. We will return to this point in section 3.2; we put off discussion until then because, as we will see, a $\mathcal{G}$-based account does not make sense of our experimental results, and in fact there are reasons to doubt that the set of alternatives that would be needed are available at all.

[10]Participants in the local group actually saw 101 trials; one non-important item appeared twice in the list of items by mistake.

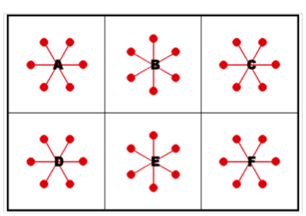
bar to proceed to the next word. After each sentence, a diagram was displayed, and participants were instructed to indicate whether the sentence was true or false as a description of the diagram (by pressing T for ‘true’ or F for ‘false’). The reading times for each word and the responses for each prompt were recorded and analysed.

Participants were recruited via Amazon Mechanical Turk and were paid for their participation. The two versions of the experiment were fielded on separate days in December 2013. 122 participants were recruited for the Global training condition and 104 for the Local condition. Both versions of the experiment contained two sets of items, the first set was presented with feedback and the second set without feedback. The items were the same across the two training conditions; only the nature of the feedback varied for a couple of items, as we describe below.

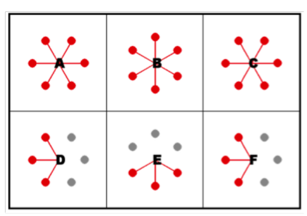
2.2.1 Items and conditions

The first set of items (with feedback) contained 48 items presented in a random order. They were of the form ‘Q1 letter is connected to Q2 of its circles’, where Q1 is either ‘every’ or ‘no’ and Q2 is either ‘some’, ‘all’ or ‘any’. Feedback was provided on the participant’s response, indicating whether or not the participant judged the sentence correctly.

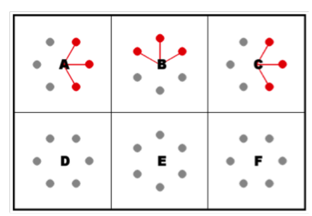
The second set of items (without feedback) contained 52 items presented in a random order (with the exception that the first two were fixed non-critical items). 28 of these items were of the same form as those in the first part of the experiment. 24 were of the form ‘X is connected to Q3 of its circles’, where ‘X’ denotes a letter (A, B, C, D, E or F) and ‘Q3’ denotes a quantifier, ‘some’, ‘none’, ‘all’, or ‘some or none’. The critical items in the second part of the experiment involved sentences of the form ‘X is connected to some of its circles’ and ‘X is connected to some or none of its circles’. Each of these sentences was presented six times to each participant, twice followed by a picture in which the relevant letter was connected to all of its circles, twice followed by a picture in which it was connected to some but not all of its circles, and twice followed by a picture in which it was connected to none of its circles. Example pictures for the letter D are shown in Figures 1a, 1b, and 1c, respectively. Control items of the form ‘X is connected to all of its circles’ and ‘X is connected to none of its circles’ were presented in the same way (one word at a time) and associated with similar displays.



(a) Example with the target letter D connected to all of its circles.



(b) Example with the target letter D connected to some but not all of its circles.



(c) Example with the target letter D connected to none of its circles.

Figure 1: Representative examples of displays, as associated for instance with the target sentence ‘D is connected to some or none of its circles.’

2.2.2 Training and feedback procedure

The Local and Global training conditions differed only with respect to the feedback that was given for the sentence ‘Every letter is connected to some of its circles’ in the first part of the experiment, as summarised in Table 1. In both training conditions, this sentence was described as ‘False’ for displays such as Figure 1a, on the grounds that every letter is in fact connected to all of its circles. In the Global condition, this sentence was described as ‘True’ for displays such as Figure 1b. However, in the Local condition, it was described as ‘False’, on the basis that some of the letters are in fact connected to all of their circles (i.e., it is not true that every letter is connected to some but not all of its circles). Each critical sentence/picture combination occurred six times. In both training conditions, the assessment of the remaining 36 items in the first phase of the experiment did not depend on any enrichment.

Description of display	Feedback for:	
	Global Group	Local Group
Every letter is connected to all of its circles (Fig. 1a).	It was false: in fact, every letter was connected to *all* of its circles.	It was false: in fact, every letter was connected to *all* of its circles.
The letters in the first row are connected to all of their circles; those in the second row to some but not all of their circles (Fig. 1b).	It was true.	It was false: in fact, the letters in the first row were connected to *all* of their circles.

Table 1: Feedback for critical items in Global and Local training conditions for the sentence ‘Every letter is connected to some of its circles.’

2.3 Results

The whole set of results and an R script that reproduces the following analyses can be found at `http://semanticsarchive.net/Archive/2M5Nzk4Z/`.

In the global training condition, we obtained 87 participants after excluding results from 4 who did not declare English as their native language, 6 for repeated Mechanical Turk IDs, and 25 for achieving less than 90% success on unambiguous items.

In the local training condition, we obtained 78 participants after excluding results from 4 participants who did not declare English as their native language, 2 participants for repeated Mechanical Turk IDs, and 20 participants for achieving less than 90% success on unambiguous items.

2.3.1 Off-line truth-value judgments

To test the effects of training and the pressure to avoid tautologies on participants’ truth-value judgments, we compared participants’ responses to the critical items in the second phase of the experiment (‘X is connected to some (or none) of its circles’). In particular, we examined the situations in which the relevant letter was connected to all of its circles, as these are the only occasions on which the response depends upon the presence or absence of the enrichment of ‘some’.

The sentences with ‘some or none’ were accepted for 20.7% of these items by Global participants, and 7.1% by Local participants. Furthermore, 26/87 Global participants (30%) gave at least one acceptance, but only 8/78 Local participants (10%) did so, representing a significant difference ($\chi^2(1) = 8.5, p = .0035$). The fact that a large majority of participants in both conditions rejected the ‘some or none’ items suggests that the pressure against tautology was generally effective in motivating the application of $\mathcal{L}$. In addition, the training regime made a significant difference, given that the Local participants were more likely to strengthen embedded instances of ‘some’ than the Global participants. This is as expected if $\mathcal{L}$ is available and preferences between $\mathcal{G}$ and $\mathcal{L}$ have been manipulated.[11]

2.3.2 Reading times

We removed from consideration any data in which any of the response times for the words in the region of interest (‘some (or none) of its circles’) exceeded 1000ms, which we took to reflect possible interruptions or loss of concentration. This excluded 9.5% of the data, leading us to analyse 477 trials from global participants and 419 trials from local participants.

As the diagrams are presented after the sentences, they can have no effect on the reading times, so we can pool data from all display conditions. Considering sentences of the form ‘X is connected to some or none of its circles’, the mean reading times (and SDs) are shown in Table 2 and presented in Figure 2a.

	some	or	none	of	its	circles
Global	317 (111)	315 (108)	346 (128)	358 (97)	350 (93)	390 (90)
Local	351 (137)	321 (111)	384 (147)	360 (107)	345 (86)	397 (109)

Table 2: Means (and SDs) of participants’ average reading times per word (in ms) for the region ‘some or none of its circles’.

To explore the effect of training on the processing of the quantifiers, we fit the reading times of the two regions (the Quantifier region ‘some or none’ and the Spillover region ‘of its circles’) with mixed models using the lme4 package (Bates et al. 2014a,b) for R (R Core Team 2015). We modeled fixed effects of region, training and their interaction. We modeled a random effect of participant, including an intercept and a random slope for region. We also modeled a random effect of item with a maximal random structure with intercept and slopes for region, training and their interaction (see Barr et al. 2013). Since the items were all identical,

[11] Turning to elementary sentences, we found sentences with ‘some’ were accepted as a true description of an ‘all’ display for 16.7% of the items by Global participants and 5.8% by Local participants; 21% of Global (18/87) and 10% of Local (8/78) participants gave at least one acceptance in this condition. This difference is surprising, given that both local and global participants should obtain the same *some but not all* meaning, albeit possibly by different means. It could be the result of a ‘no’ bias in the local group. Note however that it is not significant ($\chi^2(1) = 2.6, p = .10$), and that a ‘no’ bias cannot entirely explain the previous difference for the target condition because the two distributions, for target and elementary sentences, are different from each other ($\chi^2(7) = 141, p < 10^{-15}$).

(a) All target trials (b) Earliest target trial of each participant

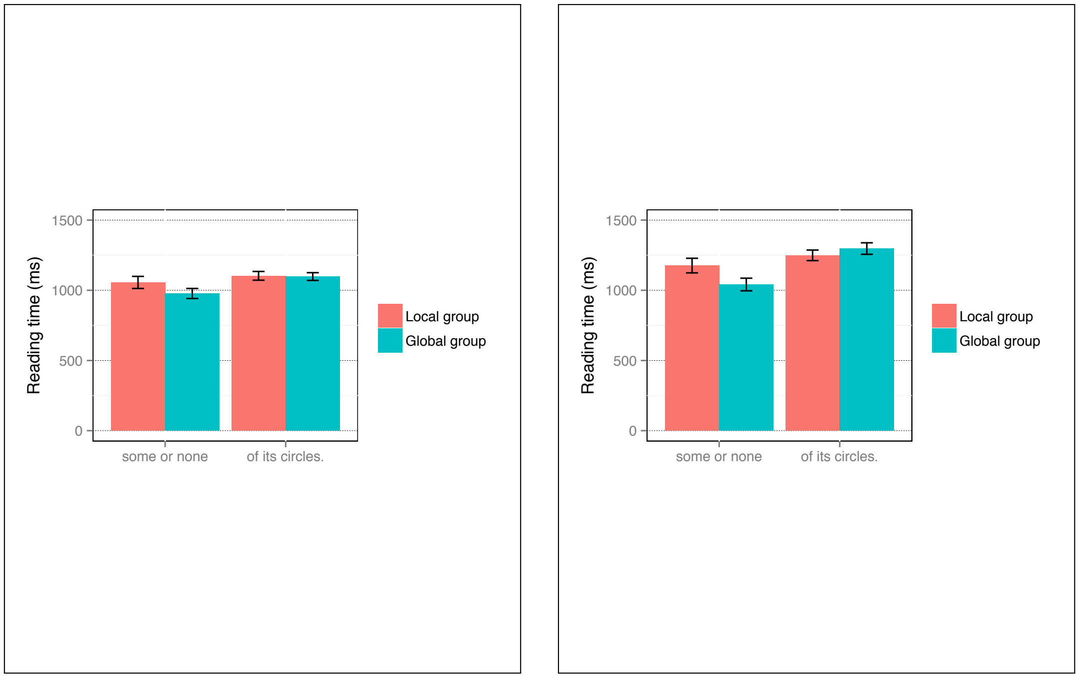


Figure 2: Reading times per regions for the critical items for (a) all trials, (b) first trials only. Error bars represent standard errors to the mean.

we encoded their identity as their position within the whole list of items of the experiment. We obtained an estimate of 79ms for the critical interaction factor, which is significantly different from zero according to a comparison with a minimally different model that simply lacks this interaction term $\chi^2(1) = 5.5, p = .019$. We note that the RTs violate the normality assumption (Shapiro-Wilk normality test: $W = .97, p < .10^{-15}$). We thus also report the results of a similar model based on log RTs (although they also violate the normality assumption, Shapiro-Wilk normality test: $W = .99, p < 10^{-10}$), which support qualitatively similar conclusions (estimate of a 6.7% effect on the RT, $\chi^2(1) = 3.9, p = .049$).[12]

We sought to explore per participant analyses only, because of the low number of items (6), because the items were repeated *identically* for a given participant, and to abstract away from the possible influence of one critical item to the next. First we ran a model as above but without per item random effects and obtained similar results ($\beta = 77$ms, $\chi^2(1) = 5.8, p = .016$, with log RTs $\beta = 6.4\%, \chi^2(1) = 3.9, p = .047$).

Strikingly, we obtain the same results by focussing the analyses on the first valid occurrence of the relevant condition for each participant ($\beta = 183$ms, $\chi^2(1) = 8.9, p = .0028$, with log RTs $\beta = 16\%, \chi^2(1) = 7.7, p = .0056$). The relevant figures are presented in Table 3 and Figure 2b.[13]

[12] We add however that the Q-Q plot of both RTs and log RTs ‘look’ reasonable, neither being clearly better or worse than the other. An anonymous reviewer prompted us to work on these issues more carefully; this and virtually all other statistical reports have been improved thanks to this reviewer’s input.

[13] For these latest two models we had to restrict the structure of the random effect to an intercept, for not having

	some	or	none	of	its	circles
Global	325 (136)	316 (119)	400 (205)	462 (200)	410 (159)	425 (136)
Local	374 (153)	318 (117)	484 (250)	423 (163)	403 (144)	423 (146)

Table 3: Means (and SDs) of reading times per word (in ms) for each participant's first item of the kind in the region 'some or none of its circles'.

We interpret these results as indicating that, relative to the Global condition, participants in the Local condition were slower to read the Quantifier region and faster to read the Spillover region. This effect was, if anything, stronger on the first occurrence of the relevant condition, when pollution by other aspects of the experimental situation are less likely to have occurred.

2.4 Discussion

Our offline results, based on truth-value judgments, provide evidence that embedded strengthening is robustly available, and can be facilitated by the need to satisfy global pragmatic pressures (recall that in both conditions the overwhelming majority of participants rejected *some or none* when 'all' is true). Furthermore, because the number of such rejections was significantly greater in the Local condition than in the Global condition, our training was effective, validating *a posteriori* the possibility of obtaining the local reading of our training *every...some...* sentence by local means.

Our online results, based on incrementally evaluated RTs, provide evidence that embedded strengthening is costly. Participants in the Local condition were slower than Global participants in the initial region of the sentence, which we interpret as indicating that local strengthening is costly (recall that our offline results show that our training was effective). The initial cost paid by local strengtheners paid dividends in later parts of the sentence: upon encountering *none*, they face no pragmatic penalty but participants who do not locally strengthen must reanalyse in order to avoid a tautologous reading. Our offline results showed that the pressure to be informative is real, and our online results show that the required reanalysis costs the Global participants with increased RTs sentence-finally. The observed mid-sentence reversal in relative RTs between Local and Global participants is a common finding in psycholinguistic studies: a decision made at one stage of evaluation will have consequences at later stages of evaluation, leading to well-known garden-path effects that sometimes can be recovered from with reanalysis (e.g., Frazier and Rayner 1982, van Gompel and Pickering 2001 and much other work). Our materials aimed to construct strengthening-based garden-path effects: those participants who don't initially strengthen save on a processing cost only to later discover that they made the wrong decision, and bear a penalty at that point.

3 General discussion

We found evidence for the application of $\mathcal{L}$ under disjunction, and we found that application of $\mathcal{L}$ is costly. This result thus extends to embedded positions the common finding that application

sufficiently many data points to do more, given that we restrict attention to one item only.

of $\mathcal{G}$ is costly. Perhaps the most natural interpretation of this extension is that $\mathcal{L} = \mathcal{G}$; that is, there is a single strengthening mechanism which comes with a processing cost in whatever position it applies. This interpretation straightforwardly follows from the grammatical theory, under which a grammatical application of exh is responsible for strengthening. Because exh can apply in global and embedded positions, we would expect it to be costly wherever it shows up. As far as we can tell, this is the simplest explanation of the current data. In the remainder of this section we would like to explore possible explanations of these results *without* appeal to exh.

3.1 Local enrichments as distorted or misguided applications of $\mathcal{G}$

At first blush, it's not clear how an embedded strengthening could apply in *some or none* with $\mathcal{G}$ alone, nor is it clear from what assumptions it would follow that such a strengthening should be costly in the way that $\mathcal{G}$ is. One possibility is that occurrences of $\mathcal{L}$ are just perverted versions of $\mathcal{G}$, which mimic $\mathcal{G}$ opportunistically, say to rescue the sentence from deviance. However, to understand the benefits of such an approach one would have to specify what the constraints are on such a perversion of $\mathcal{G}$ into $\mathcal{L}$, and how the output of such an $\mathcal{L}$ is computed (if not by manipulating assumptions about the speaker's beliefs, as for $\mathcal{G}$).

Another possibility is that participants might strengthen using $\mathcal{G}$ together with the default assumption that the continuations of the sentence are all elementary.[14] Thus, upon reading *some* a participant might guess that the sentence will be elementary, *some X Y* (for some X and Y), and then compute a global enrichment over such continuations: within some reasonable limits for the values of X and Y, the sentence *some X Y* will be entailed by *all X Y*. This could justify the incremental computation of a strengthened meaning for *some*, but crucially there is no appeal to $\mathcal{L}$. This line of reasoning is not complete, however. Note that upon encountering *or*, participants reasoning in this way would have to incorporate their (mistaken) computation with the final semantic decoding of the sentence in some way that is left undefined. Most plausibly, they would have to plainly revise their analysis: the connective provides evidence that the sentence is complex, which in turn would 'cancel' the strengthened meaning of *some* ($\mathcal{G}$ does not have access to sub-constituents; cf. footnote 3). The prediction, then, is that a cost earlier in the sentence (corresponding to enrichment) would lead to an additional cost after processing *or* (corresponding to cancellation and revised computation). But this is the opposite of what we find.

3.2 Seemingly local effects via the manipulation of alternative sets

In this section we entertain (and rule out) the possibility that the seemingly local readings can be derived by global means, provided the right set of alternatives can be given to them. This discussion is therefore highly sensitive to the theory of alternatives one chooses to follow.

It is useful to recall from section 1.3 that, in cases like *every...some*, an apparent embedded strengthening can be derived with global reasoning alone together with certain assumptions about alternatives. Thus, what we have been referring to as 'local' vs. 'global' training might

[14] The default assumption could be motivated by a parsing strategy under which participants analyze the sentence so as to find the least complex parse consistent with the overt string (cf. Miller and Chomsky 1963 and much work since).

instead have been training participants to select one set of alternatives instead of another. Recall that the 'local' reading 'every letter is connected to some but not all of its circles' can be derived with $\mathcal{G}$ if alternatives may be generated with multiple scalar replacements, and in particular if the replacements lead to an alternative that does not entail the asserted sentence. This assumption allows *some letter is connected to all of its circles* to be an alternative, whose negation – in conjunction with the assertion – is equivalent to the local reading. Thus, our Local participants may have been trained to generate all non-weaker alternatives using multiple replacements, whereas Global participants might have been trained to use a more restricted set of alternatives that prevents *some...all* from being an alternative.[15]

Fox (2007a) and (Magri 2009a, Part I) propose a natural constraint on alternatives that provides the intended restriction. This constraint, call it F&M's constraint, proposes that alternatives are generated step-wise through a sequence of replacements such that each replacement operation starting from an alternative A must not lead to an alternative A' that is entailed by the starting point A. Going from *every...some* to *some...all* requires two replacements: *every* in subject position must be replaced by *some* and *some* in object position must be replaced by *all*. However, there is no way to make these replacements in sequence while satisfying F&M's constraint (assuming existential import of quantifiers): the replacement of *every* with *some* in subject position will create a conflict with the constraint, no matter when it occurs in the sequence of replacements. If we start by replacing the object quantifier, the computation fails at the second step because *some...all* is weaker than *every...all*, and the computation fails if we instead start with the subject replacement, because *some...some* is weaker than *every...some*. Thus, the only alternative that is available under this constraint is *every...all*, and thus only the 'global' reading can be generated.

Suppose, then, that there is no local strengthening mechanism $\mathcal{L}$, and suppose that our 'Local' participants differed from our 'Global' participants in the choice of whether they followed F&M's constraint in generating the alternatives used by $\mathcal{G}$: the Local group ignored F&M's constraint while the Global group obeyed it. Under this assumption, assuming that alternatives are derived by replacing nodes with other lexical items or with subconstituents (Katzir 2007), we obtain distinct sets of alternatives for our target sentence *some or none* too. The alternatives for the Local group are $C_L = \{\exists, \neg\exists, \forall, \forall \vee \neg\exists, \neg\forall, \exists \wedge \neg\forall\}$ and the alternatives for the Global group are $C_G = \{\exists, \neg\exists, \forall, \forall \vee \neg\exists\} = C_L \setminus \{\neg\forall, \exists \wedge \neg\forall\}$. The difference between C_G and C_L is that F&M's constraint prevents $\neg\forall$ and $\exists \wedge \neg\forall$ from being generated. If $\mathcal{G}$ is identified with the standard innocent exclusion algorithm for strengthening (Sauerland 2004, Fox 2007a), which we assume to be familiar, $\mathcal{G}$ is vacuous when applied to our target sentence (*some or none*) with C_L as alternatives, but when applied with C_G as alternatives it outputs $\neg\forall$.[16] Thus, under these assumptions about alternatives, which crucially are motivated by independent considerations, it *is* possible to get the desired $\neg\forall$ strengthened meaning out of $\exists \vee \neg\exists$ using $\mathcal{G}$ alone.[17]

[15]Crucially, given that strengthening is known to occur in non-monotonic environments (see e.g., Chemla and Spector 2011 for experimental data), a limitation to only *stronger* alternatives will not suffice.

[16]Note that each alternative in C_L has its negation as an alternative, which means each alternative has a symmetric alternative (cf. note 1) and hence none is 'innocently excludable' (Fox 2007a). On the other hand, the maximal consistent exclusions of C_G are $\{\exists, \forall\}$ and $\{\forall, \neg\exists, \forall \vee \neg\exists\}$; their intersection is $\forall$, and thus $\forall$ is innocently excludable.

[17]We thank an anonymous reviewer, Danny Fox and Bernhard Schwarz for urging us to investigate this possibility more carefully, and for pointing out an error in an earlier version.

Although we have a purely global mechanism for avoiding tautology in *some or none*, this mechanism makes strange predictions about the offline data. Specifically, the global group is trained to go with C_G, and thus to avoid tautology. The obvious prediction, then, is that we should find *more* rejections of *some or none* when ‘all’ is true for participants in the Global group than for participants in the Local group, who, at best, need to revise their original assumption that the relevant set of alternative is C_L. But this prediction is the opposite of what we find.

Let us explain the situation in more abstract terms. F&M’s constraint filters alternatives. Quite often, more alternatives lead to more inferences, simply because there are more *candidates* for being excludable. This is the situation in the training case (*every...some...*), where the potential *some...every...* alternative is innocently excludable, provided that it is a legitimate alternative. But in the target case (*some or none*) the opposite actually happens: *fewer* alternatives lead to more inferences. This is so because, in this particular case, the biggest subset of alternatives is full of symmetry (i.e., conflict), which prevents alternatives from being innocently excludable (for discussion of possibly similar cases, see e.g., Fox and Hackl 2006, Fox 2007b, Singh 2008a,b, Katzir 2013, Trinh and Haida 2015). Somewhat counterintuitively then, this is a case where a smaller subset of alternatives gives rise to a bigger set of innocently excludable alternatives. Hence F&M’s constraint puts opposite pressure on the training and target cases. What we find however is a parallelism between them, whereby stronger inferences in training lead to more enrichments in targets, too.

In sum, even without considering the necessary assumptions about processing concerning $\mathcal{G}$ that could explain the reading time differences between our two groups, the possibility of a global derivation of the non-tautologous reading of our target sentence is hard to reconcile with the current set of data. Instead, we find evidence in favor of a common mechanism that leads to strengthening of the training sentences and avoids the tautologous reading for the target sentences. Local $\mathcal{L}$ strengthening may therefore be at play not only for the target sentences but also for the training, controversial *every...some...* sentences.[18]

3.3 Conclusion

We presented a judgment and reading time study of sentences containing the scalar item *some* in embedded positions. The resulting data are hard to reconcile with approaches of scalar implicatures based on global $\mathcal{G}$ mechanisms only, but can be explained if we assume: (i) that local and global mechanisms are identical ($\mathcal{L} = \mathcal{G}$), and (ii) that application of this mechanism is costly, wherever it appears.

References

Márta Abrusán. *Weak Island Semantics*. Oxford University Press, Oxford, 2014.

[18] An anonymous reviewer suggests that $\mathcal{G}$ could produce the desired $\neg\forall$ reading if the two disjuncts $\exists$ and $\neg\exists$ were the only alternatives to the sentence. This is correct, but we don’t quite see how to connect this possibility with our training. Furthermore, there are independent considerations that would rule out such a set of alternatives; see Trinh and Haida (2015) (thanks to another reviewer for pointing this out to us).

Dale J Barr, Roger Levy, Christoph Scheepers, and Harry J Tily. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of memory and language*, 68(3): 255–278, 2013.

Douglas Bates, Martin Maechler, Ben Bolker, and Steven Walker. *lme4: Linear mixed-effects models using Eigen and S4*, 2014a. URL `http://CRAN.R-project.org/package=lme4`. R package version 1.1-7.

Douglas Bates, Martin Maechler, Benjamin M. Bolker, and Steven Walker. lme4: Linear mixed-effects models using eigen and s4, 2014b. URL `http://arxiv.org/abs/1406.5823`. ArXiv e-print; submitted to *Journal of Statistical Software*.

Lewis Bott and Ira Noveck. Some utterances are underinformative: The onset and time course of scalar inferences. *Journal of Memory and Language*, 53:437–457, 2004.

Lewis Bott, Todd M. Bailey, and Daniel Grodner. Distinguishing speed from accuracy in scalar implicatures. *Journal of Memory and Language*, 66:123–142, 2012.

Richard Breheny, Napoleon Katsos, and John Williams. Are generalized scalar implicatures generated by default? An on-line investigation into the role of context in generating pragmatic inferences. *Cognition*, 100:434–463, 2006.

Emmanuel Chemla. Similarity: Towards a unified account of scalar implicatures, free choice permission, and presupposition projection. Ms. CNRS, ENS, LSCP Paris, 2009.

Emmanuel Chemla. Frequencies of apparent Hurford constraint obviations track implication rates. Ms. CNRS, ENS, LSCP Paris, 2013.

Emmanuel Chemla and Lewis Bott. Processing inferences at the semantics/pragmatics frontier: Disjunctions and free choice. *Cognition*, 130:380–396, 2014.

Emmanuel Chemla and Raj Singh. Remarks on the experimental turn in the study of scalar implicature, Part I. *Language and Linguistics Compass*, 8(9):373–386, 2014a.

Emmanuel Chemla and Raj Singh. Remarks on the experimental turn in the study of scalar implicature, Part II. *Language and Linguistics Compass*, 8(9):387–399, 2014b.

Emmanuel Chemla and Benjamin Spector. Experimental evidence for embedded scalar implicatures. *Journal of Semantics*, 28:359–400, 2011.

Gennaro Chierchia. Scalar implicatures, polarity phenomena, and the syntax-pragmatics interface. In A. Belletti, editor, *Structures and Beyond*. Oxford University Press, Oxford, 2004.

Gennaro Chierchia. Broaden your views: Implicatures of domain widening and the ‘logicality’ of language. *Linguistic Inquiry*, 37(4):535–590, 2006.

Gennaro Chierchia, Danny Fox, and Benjamin Spector. The grammatical view of scalar implicatures and the relationship between semantics and pragmatics. In Paul Portner, Claudia Maienborn, and Klaus von Heusinger, editors, *Handbook of Semantics*. Mouton de Gruyter, New York, NY, 2008.

Charles Clifton and Chad Dube. Embedded implicatures observed: A comment on Geurts and Pouscoulous. *Semantics and Pragmatics*, 3(7):1–13, 2010.

Kai von Fintel. Exceptive constructions. *Natural Language Semantics*, 1:123–148, 1993.

Kai von Fintel and Irene Heim. Notes on implicature. Lecture Notes, 24.954: Pragmatics in Linguistic Theory, MIT, 1999.

Danny Fox. *Economy and Semantic Interpretation*. MITWPL and MIT Press, Cambridge, MA, 2000.

Danny Fox. Scalar implicatures: Pragmatics or grammar or both. Lecture Notes, 24.959: Topics in Semantics, MIT, 2006.

Danny Fox. Free choice disjunction and the theory of scalar implicature. In Uli Sauerland and Penka Stateva, editors, *Presupposition and Implicature in Compositional Semantics*, pages 71–120. Palgrave Macmillan, New York, NY, 2007a.

Danny Fox. Too many alternatives: Density, symmetry, and other predicaments. *Proceedings of SALT 17*, 2007b.

Danny Fox. Cancelling the Maxim of Quantity: another challenge for a Gricean theory of Scalar Implicatures. In press, *Semantics and Pragmatics*, 2013.

Danny Fox and Martin Hackl. The universal density of measurement. *Linguistics and Philosophy*, 29:537–586, 2006.

Danny Fox and Roni Katzir. On the characterization of alternatives. *Natural Language Semantics*, 19:87–107, 2011.

Danny Fox and Benjamin Spector. Economy and embedded exhaustification. Handout, MIT and Harvard University/ENS, 2008.

Danny Fox and Benjamin Spector. Economy and embedded exhaustification. Accepted for publication in *Natural Language Semantics*, 2015.

Michael Franke. Quantity implicatures, exhaustive interpretation, and rational conversation. *Semantics and Pragmatics*, 4(1):1–82, June 2011. doi: 10.3765/sp.4.1.

Lyn Frazier and Keith Rayner. Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology*, 14:178–210, 1982.

Jon Gajewski. On analyticity in natural language. Manuscript, MIT, 2004.

Jon Gajewski and Yael Sharvit. In defense of the grammatical approach to scalar implicature. *Natural Language Semantics*, 20(1):31–57, 2012.

L. T. F. Gamut. *Logic, Language, and Meaning*, volume 1. University of Chicago Press, Chicago, IL, 1991.

Gerald Gazdar. *Pragmatics*. Academic Press, New York, NY, 1979.

Bart Geurts. Scalar implicature and local pragmatics. *Mind and Language*, 24:51–79, 2009.

Bart Geurts and Nausicaa Pouscoulous. Embedded implicatures?!? *Semantics and Pragmatics*, 2:1–34, 2009.

Bart Geurts and Bob van Tiel. Embedded scalars. *Semantics and Pragmatics*, 6(9):1–37, 2013.

H.P. Grice. Logic and conversation. William James Lectures, Harvard University, 1967.

Jeroen Groenendijk and Martin Stokhof. Studies on the semantics of questions and the pragmatics of answers. Doctoral Dissertation, University of Amsterdam, 1984.

Irene Heim. Predicates or formulas? Evidence from VP-ellipsis. In Aaron Lawon and Eund Cho, editors, *Proceedings of SALT 7*, pages 197–221. CLC Publications, Ithica, NY, 1997.

Laurence R. Horn. On the semantic properties of logical operators in english. Doctoral Dissertation, UCLA, 1972.

Yi Ting Huang and Jesse Snedeker. Online interpretation of scalar quantifiers: Insight into the semantics-pragmatics interface. *Cognitive Pyschology*, 58:376–415, 2009.

James Hurford. Exclusive or inclusive disjunction. *Foundations of Language*, 11:409–411, 1974.

Napoleon Katsos and Chris Cummins. Pragmatics: From theory to experiment and back again. *Language and Linguistics Compass*, 4:282–295, 2010.

Roni Katzir. Structurally defined alternatives. *Linguistics and Philosophy*, 30:669–690, 2007.

Roni Katzir. A note on contrast. *Natural Language Semantics*, 21(4):333–343, 2013.

Roni Katzir and Raj Singh. Hurford disjunctions: Embedded exhaustification and structural economy. In Urtzi Etxeberria, Anamaria Falaus, Aritz Irurtzun, and Bryan Leferman, editors, *Proceedings of* Sinn und Bedeutung 18, pages 201–216, 2014. In press, Proceedings of *Sinn und Bedeutung 18*.

Christopher Kennedy. Predicates *and* formulas: Evidence from ellipsis. In Luka Crnic and Uli Sauerland, editors, *The Art and Craft of Semantics: A Festschrift for Irene Heim*, volume 1, pages 253–277. MITWPL and MIT Press, Cambridge, MA, 2014.

Anthony Kroch. Lexical and inferred meanings for some time adverbs. *Quarterly Progress Reports of the Research Laboratory of Electronics, MIT*, 104:260–267, 1972.

David Lewis. Relevant implication. *Theoria*, 54:161–174, 1988.

Giorgio Magri. *A theory of individual-level predicates based on blind mandatory scalar implicatures. Constraint promotion for Optimality Theory*. PhD thesis, MIT, 2009a.

Giorgio Magri. A theory of individual level predicates based on blind mandatory scalar implicatures. *Natural Language Semantics*, 17:245–297, 2009b.

Giorgio Magri. Another argument for embedded scalar implicatures based on oddness in downward entailing environments. *Semantics and Pragmatics*, 4(6):1–51, 2011.

Paul Marty, Emmanuel Chemla, and Benjamin Spector. Interpreting numerals and scalar items under memory load. *Lingua*, 133:152–163, 2013.

Clemens Mayr and Jacopo Romoli. Redundancy, exhaustification, and the notion of local context. Manuscript, 2014.

James D. McCawley. *Everything Linguists have Always Wanted to Know about Logic*. University of Chicago Press, Chicago, 1981.

Marie-Christine Meyer. Generalized free choice and missing alternatives. Manuscript, MIT, 2013.

George Miller and Noam Chomsky. Finitary models of language users. In R.D. Luce, R.R. Bush, and E. Galanter, editors, *Handbook of Mathematical Psychology*, volume 2, pages 419–491. Wiley, New York, NY, 1963.

Ira Noveck and Andres Posada. Characterizing the time course of an implicature: An evoked potentials study. *Brain and Language*, 85(2):203–210, 2003.

Ira Noveck and Anne Reboul. Experimental pragmatics: A Gricean turn in the study of language. *Trends in Cognitive Sciences*, 12:425–431, 2008.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2015. URL `http://www.R-project.org/`.

Tanya Reinhart. *Interface Strategies*. MIT Press, Cambridge, MA, 2006.

Craige Roberts. Information structure and discourse: Towards an integrated formal theory of pragmatics. In J.H. Yoon and Andreas Kathol, editors, *OSU Working Papers in Linguistics*, pages 91–136. 1996.

Robert van Rooij and Katrin Schulz. Exhaustive interpretation of complex sentences. *Journal of Logic, Language, and Information*, 13:491–519, 2004.

Benjamin Russell. Against grammatical computation of scalar implicatures. *Journal of Semantics*, 23:361–382, 2006.

Uli Sauerland. Scalar implicatures in complex sentences. *Linguistics and Philosophy*, 27: 367–391, 2004.

Uli Sauerland. The computation of scalar implicatures: pragmatic, lexical, or grammatical? *Language and Linguistics Compass*, 6(1):36–49, 2012.

Katrin Schulz and Robert van Rooij. Pragmatic meaning and non-monotonic reasoning: The case of exhaustive interpretation. *Linguistics and Philosophy*, 29:205–250, 2006.

Mandy Simons. *Issues in the Semantics and Pragmatics of Disjunction*. Garland, New York, NY, 2000.

Raj Singh. On the interpretation of disjunction: Asymmetric, incremental, and eager for inconsistency. *Linguistics and Philosophy*, 31:245–260, 2008a.

Raj Singh. Modularity and locality in interpretation. Doctoral Dissertation, MIT, 2008b.

Raj Singh. Implicature cancellation and exhaustivity. Presented at *SALT 22*, 2012.

Benjamin Spector. Scalar implicatures: Exhaustivity and Gricean reasoning. In Maria Aloni, Alastair Butler, and Paul Dekker, editors, *Questions in Dynamic Semantics*. Elsevier, Amsterdam, 2005.

Benjamin Spector. Aspects de la pragmatique des opérateurs logiques. Doctoral Dissertation, Université Paris 7, 2006.

Robert Stalnaker. Assertion. In Peter Cole, editor, *Pragmatics*, pages 315–332. Academic Press, New York, NY, 1978.

Bob van Tiel, Emiel van Miltenburg, Natalia Zevakhina, and Bart Geurts. Scalar diversity. *Journal of Semantics*, in press.

Tue Trinh and Andreas Haida. Constraining the derivation of alternatives. *Natural Language Semantics*, pages 1–22, 2015. ISSN 0925-854X. doi: 10.1007/s11050-015-9115-y. URL `http://dx.doi.org/10.1007/s11050-015-9115-y`.

Roger P.G. van Gompel and Martin J. Pickering. Reanalysis in sentence processing: Evidence against current constraint-based and two-stage models. *Journal of Memory and Language*, 45:225–258, 2001.